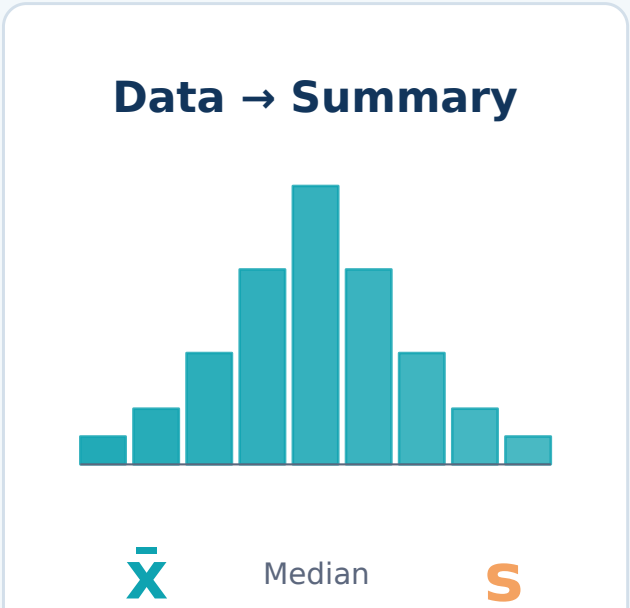
Descriptive Statistics

Mean, Median, Mode and Standard Deviation

A student-friendly lecture on summarizing, comparing, and interpreting quantitative data in research.

Author: Muhammad Hassan
Website: ResearchMethod.net

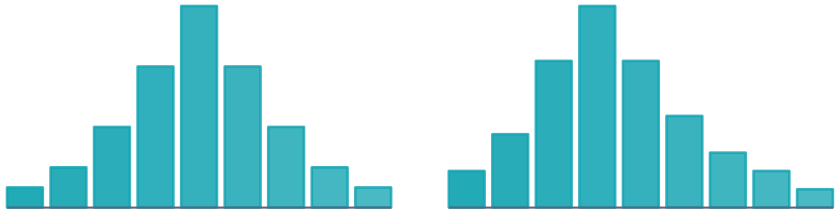


Lecture Overview

Raw data | Clean data | Summaries | Interpretation | Reporting

What descriptive statistics are and why they matter
How to calculate and interpret mean, median, mode, and standard deviation
How distribution shape, outliers, and measurement level affect interpretation
How to report descriptive statistics ethically and transparently
Practice activities, knowledge check, and answer key

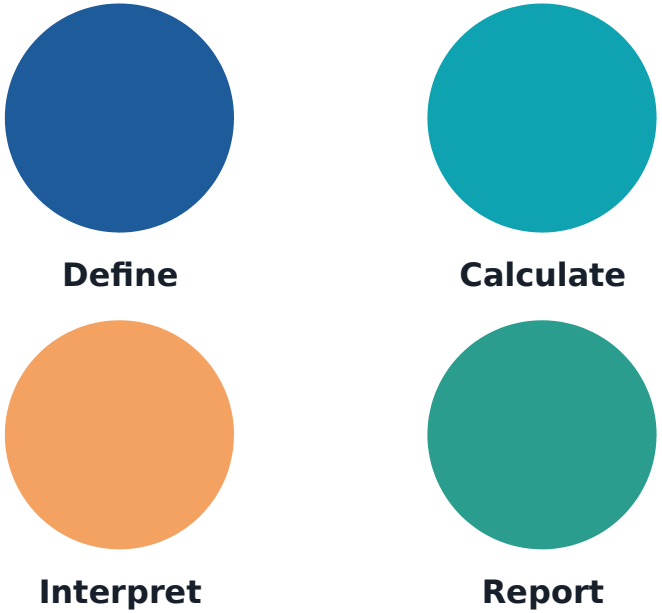
Descriptive statistics answer questions such as: What is typical? How much do scores vary? Which value occurs most often?



Two different data shapes can have similar averages but very different meanings.

Learning Objectives

Define descriptive statistics and distinguish them from inferential statistics.
Calculate the mean, median, mode, range, variance, and standard deviation for simple datasets.
Explain what each statistic communicates about a dataset.
Choose appropriate measures of central tendency and variability for different data types.
Interpret the effect of outliers and skewed distributions on summary statistics.
Report descriptive statistics clearly, ethically, and in academic style.

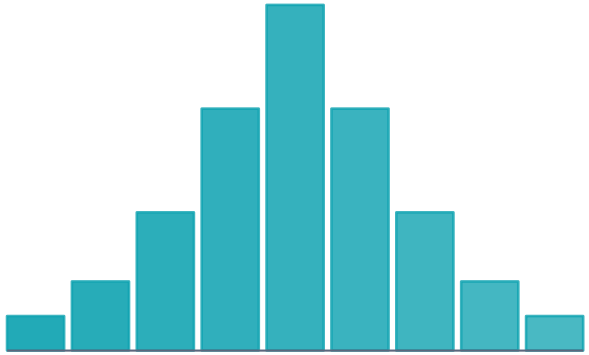


Measurable skills for research data analysis

What Are Descriptive Statistics?

Descriptive statistics are numerical and visual summaries used to organize, simplify, and communicate the main features of a dataset.

They describe what the collected data show.
They do not test hypotheses or prove cause-and-effect relationships.
They help researchers understand patterns before choosing advanced statistical tests.
They make tables, questionnaires, exam scores, survey responses, and measurements easier to interpret.



Example: A researcher can summarize 200 exam scores using an average score and standard deviation instead of listing every score.

Key Definitions

Term	Student-friendly meaning
Dataset	A collection of values gathered from participants, documents, records, or observations.
Variable	A characteristic that can vary, such as age, score, income, response time, or satisfaction.
Observation	One recorded value or case in a dataset.
Central tendency	The “typical” or central value of a dataset: mean, median, and mode.
Variability	How spread out the data values are: range, variance, and standard deviation.
Outlier	A value that is unusually far from most other values.

Background and Academic Context

Descriptive statistics are part of the broader tradition of data summarization and exploratory data analysis.
In research, they usually appear before inferential tests in the results section.
They are used across disciplines because most studies need to describe participants, measurements, and patterns.
Modern statistical software calculates these values quickly, but researchers still need to understand their meaning.

Good descriptive analysis is not just a calculation step; it is a thinking step that helps researchers decide what the data can and cannot support.

Main Concepts: Center and Spread

A useful description of quantitative data usually needs two questions: What is typical? and How much do values differ?

Measures of Center

Mean
Median
Mode

Tells us where the data are located.

Measures of Spread

Range
Variance
Standard deviation

Tells us how varied or dispersed the data are.

Types of Data and Measurement Levels

Level	Meaning	Common summaries
Nominal	Categories with no rank, e.g., department, gender category, software type	Mode, frequencies
Ordinal	Ranked categories, e.g., satisfaction level, agreement scale	Median, mode, frequencies
Interval	Equal intervals but no true zero, e.g., temperature in °C	Mean, median, standard deviation
Ratio	Equal intervals and true zero, e.g., age, income, time, score counts	Mean, median, standard deviation

Choose statistics that match the type of data. For example, a mean is usually not meaningful for purely nominal categories.

Mean: The Arithmetic Average

Mean = Σx / n

Add all values and divide by the number of values.

Exam scores	70	75	80	85	90
Calculation	70+75+80+85+90 = 400			**400/5 = 80**	

Interpretation: The average exam score is 80. This suggests the central performance level, but it does not show how spread out the scores are.

Best for interval or ratio data.
Uses every value in the dataset.
Sensitive to extreme values and outliers.
Often reported with standard deviation.

Median: The Middle Value

Median = middle ordered value

Arrange values from smallest to largest, then locate the center.

Ordered values

Median = 71

Useful when data are skewed or contain outliers.
Suitable for ordinal, interval, and ratio data.
For an odd number of values, it is the single middle value.
For an even number of values, average the two middle values.

When one value is extreme, the median often gives a more stable picture of the typical case than the mean.

Mode: The Most Frequent Value

Mode = value with highest frequency

A dataset may have one mode, more than one mode, or no repeated mode.

Works for nominal, ordinal, interval, and ratio data.
Useful for identifying the most common category or response.
Can be unimodal, bimodal, multimodal, or have no mode.
Does not use all values and may be unstable in small samples.

Response	Frequency
Agree	12
Neutral	8
Disagree	5
Strongly agree	15

Mode = Strongly agree, because it appears most often.

Mean, Median, and Mode Compared

Measure	Best used when…	Main caution
Mean	Data are numerical and roughly symmetric.	Can be pulled by outliers.
Median	Data are skewed, ordinal, or contain extreme values.	Ignores the exact size of most values.
Mode	Data are categorical or the most common value matters.	May be unclear when several values tie.

A good researcher asks: “Which summary gives the most honest picture of this variable?”

Variability: Why Spread Matters

Two groups can have the same mean but very different levels of consistency.

Class A: scores close together

Class B: scores spread out

Variability describes how different the values are from each other.
Low variability means values are relatively consistent.
High variability means values are widely dispersed.
Standard deviation is one of the most common measures of variability.

Range, Variance, and Standard Deviation

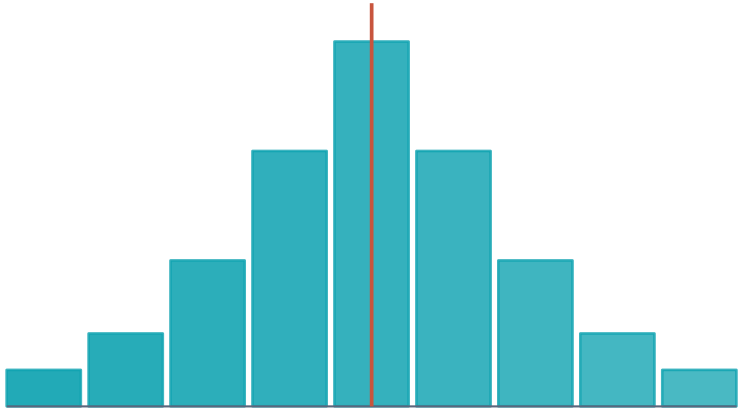
Measure	Definition	Simple interpretation
Range	Largest value − smallest value	Total width of the dataset
Variance	Average squared distance from the mean	How much values vary, in squared units
Standard deviation	Square root of variance	Typical distance from the mean, in original units

Sample SD: $s = \sqrt{[\Sigma(x - \bar{x})^2 / (n - 1)]}$

The sample standard deviation estimates spread in a population using sample data.

Interpreting Standard Deviation

A small standard deviation means values are close to the mean.
A large standard deviation means values are more spread out.
Standard deviation is affected by extreme values.
It is most meaningful for numerical variables measured on interval or ratio scales.
When data are approximately normal, many values fall within one or two standard deviations of the mean.



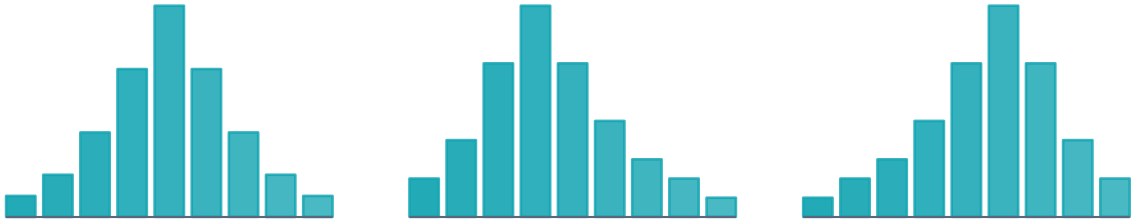
Typical distance from the mean

Report SD with the mean: “M = 72.4, SD = 8.6.” This communicates both typical value and variation.

Distribution Shape and Outliers

Distribution shape affects how descriptive statistics should be interpreted:

- Symmetric data: mean and median are often close.
- Right-skewed data: mean is often higher than the median.
- Left-skewed data: mean is often lower than the median.
- Outliers can strongly influence the mean and standard deviation.



Symmetric **Right-skewed** **Left-skewed**

Rule of thumb: when distributions are strongly skewed, report the median and interquartile range alongside or instead of the mean and standard deviation.

Choosing the Right Descriptive Statistic

Identify variable type | **Inspect distribution** | **Check outliers** | **Select measures** | **Report transparently**

Situation	Recommended summary
Nominal categories	Frequencies, percentages, mode
Ordinal ratings	Median, mode, frequencies; sometimes mean for scale scores
Symmetric numerical data	Mean and standard deviation
Skewed numerical data	Median and interquartile range; report outliers if important
Small dataset with extreme values	Median, range, individual values when appropriate

Step-by-Step Process for Descriptive Analysis

Process

1. Define variables
2. Clean data
3. Choose summaries
4. Calculate values
5. Interpret
6. Report

1. Confirm what each variable measures and its level of measurement.
2. Screen for missing values, impossible values, and potential outliers.
3. Select central tendency and variability measures that fit the variable.
4. Calculate statistics manually or with software and verify key results.
5. Interpret values in relation to the research question.
6. Report clearly with sample size, units, and relevant notes.

Worked Example: Small Dataset

Dataset: weekly study hours for 8 students

Student	A	B	C	D	E	F	G	H
Hours	4	6	6	7	8	9	9	15

Mean = 64 / 8 = 8

Average weekly study hours.

Median = (7 + 8) / 2 = 7.5

Middle value after ordering.

Mode = 6 and 9

Two values are tied for most frequent.

Interpretation: The mean is slightly higher than the median because one student studied 15 hours.

Practical Examples Across Disciplines

Discipline	Variable	Useful descriptive statistics
Social sciences	Monthly household income	Median, mean, standard deviation, range
Business	Customer satisfaction score	Mean, median, mode, standard deviation
Education	Exam performance	Mean, median, standard deviation
Health sciences	Blood pressure reading	Mean, standard deviation, minimum, maximum
Psychology	Anxiety scale score	Mean, standard deviation, reliability context
Information technology	System response time	Median, mean, standard deviation, outliers

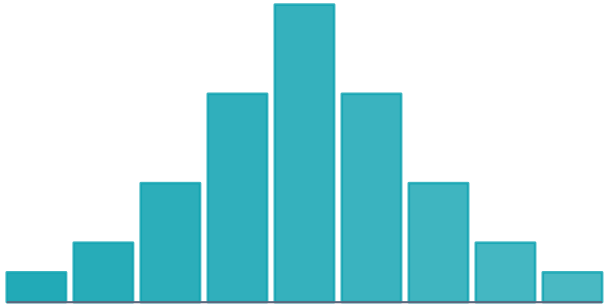
Research Example: Student Sleep and Academic Performance

Scenario: A university researcher wants to describe the sleep patterns and GPA distribution of undergraduate students.

Element	Example
Research topic	Sleep duration and academic performance among university students
Research problem	Students report irregular sleep, but the pattern and variation are unclear.
Research question	What are the average sleep hours and GPA levels of students in the sample?
Participants	Undergraduate students from multiple departments
Data collection	Short survey asking typical sleep hours and self-reported GPA
Analysis method	Mean, median, mode, standard deviation, range, and distribution inspection
Expected interpretation	Summaries describe typical sleep duration, variation, and whether extreme values affect the mean.

Advantages of Descriptive Statistics

Simple to calculate and easy to communicate.
Useful for summarizing large datasets in a compact format.
Helps researchers identify trends, patterns, and unusual values.
Supports data screening before advanced statistical tests.
Useful for comparing groups or variables at an initial level.
Makes research findings more accessible to readers.

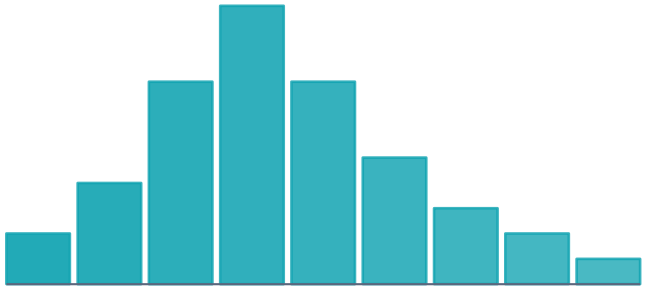


Strength: Descriptive statistics turn many values into a meaningful research summary without overclaiming.

Limitations and Challenges

They describe data but do not explain why patterns occur.
They cannot establish causal relationships.
A single number can oversimplify a complex dataset.
Outliers and skewed distributions can distort some measures.
Small samples may produce unstable summaries.
Poor data quality leads to misleading summaries.

Limitation example: The mean income of a group may look high because of one very wealthy participant; the median may better describe the typical participant.



Common Mistakes to Avoid

Checklist question: Would another researcher understand how this number was calculated and what it represents?

Reporting only the mean and ignoring variability.
Using the mean for categories where it has no meaning.
Forgetting to mention missing data or sample size.
Confusing standard deviation with standard error.
Deleting outliers without a defensible reason.
Interpreting descriptive statistics as proof of causation.
Rounding inconsistently or reporting excessive decimal places.

Best Practices for Reporting Descriptive Statistics

Report sample size clearly for each variable or group.
Use mean and standard deviation for roughly symmetric numerical data.
Use median and range or interquartile range when data are skewed.
Use frequencies and percentages for categorical variables.
Include units of measurement, such as years, seconds, or scale points.
Inspect distributions visually before interpreting statistics.
Explain any data cleaning, missing data, or outlier decisions.

Example report: “Students studied an average of 8.0 hours per week (SD = 3.4, n = 8). The median was 7.5 hours.”

Use consistent decimal places and avoid false precision.

Comparison Table: Descriptive Statistics at a Glance

Statistic	Answers	Good for	Weakness
Mean	What is the average?	Symmetric numerical data	Sensitive to outliers
Median	What is the middle?	Skewed or ordinal data	Uses less information
Mode	What is most common?	Categorical or repeated values	May be multiple or absent
Standard deviation	How spread out are values?	Numerical data with mean	Affected by outliers

Use the table as a quick decision guide when planning descriptive analysis for a research project.

Ethical Considerations in Descriptive Reporting

Ethical reporting means the summary should help readers understand the data, not persuade them through distortion.

Protect confidentiality when reporting small groups or sensitive characteristics.
Avoid selective reporting that hides inconvenient patterns.
Explain missing data and outlier handling transparently.
Use truthful labels, units, and scales in tables and charts.
Do not manipulate axis scales or categories to exaggerate effects.
Report enough information for readers to understand the data context.

Classroom Activity: Summarize a Dataset

Activity objective

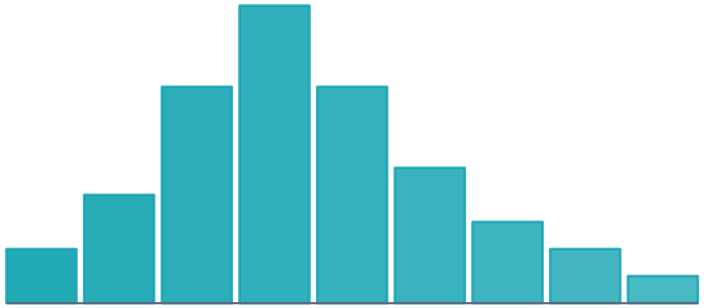
Practice calculating and interpreting measures of center and spread.

Instructions

Work individually or in groups of 3–4.
Use the dataset: 5, 6, 6, 7, 8, 9, 9, 15.
Calculate mean, median, mode, range, and standard deviation.
Decide which statistic best represents the typical value.
Write a two-sentence interpretation.

Expected outcome

Students should recognize that the value 15 increases the mean and spread, while the median provides a more resistant measure of typical study hours.



Discussion Questions

Classroom Discussion

When is the mean a misleading measure of a typical value?
Why is the median useful for income or response-time data?
How can a large standard deviation change the interpretation of a mean?
Should outliers ever be removed from a dataset? Why or why not?
Why do categorical variables require different summaries from numerical variables?
How can descriptive statistics be used ethically in research reporting?
What information should accompany a reported mean in a journal article?

Knowledge Check

1. Which measure is most affected by an extreme outlier?
2. What is the median of 4, 6, 9, 10, 12?
3. Which measure is best for the most common category?
4. What does standard deviation describe?
5. Should the mean be used for nominal categories? Explain briefly.
6. For right-skewed income data, which measure of center is often useful?
7. What two statistics are commonly reported together for symmetric numerical data?
8. Name one ethical issue in reporting descriptive statistics.

Knowledge Check: Answer Key

1. Mean.
2. 9.
3. Mode.
4. How spread out values are, or the typical distance from the mean.
5. Usually no; nominal categories are labels, so frequencies and mode are more meaningful.
6. Median.
7. Mean and standard deviation.
8. Examples: hiding missing data, deleting outliers without justification, misleading charts, or revealing identities in small groups.

Key Takeaways

Descriptive statistics summarize and communicate the main features of data.
Mean, median, and mode describe central tendency in different ways.
Standard deviation describes variability and should often accompany the mean.
Distribution shape and outliers influence which summaries are most appropriate.
Good descriptive analysis requires interpretation, not just calculation.
Ethical reporting requires transparency, context, and honesty about limitations.

Best summary = correct statistic + correct context + clear interpretation.

Conclusion

Descriptive statistics provide the foundation for understanding quantitative research data. They help researchers identify typical values, variation, patterns, and potential problems before moving to deeper analysis.

Use the mean, median, and mode to describe central tendency.
Use standard deviation and related measures to describe spread.
Select statistics based on data type, distribution shape, and research purpose.
Report summaries clearly so readers can evaluate the evidence.

References

APA 7th

Agresti, A., & Franklin, C. A. (2018). Statistics: The art and science of learning from data (4th ed.). Pearson.
Field, A. (2018). Discovering statistics using IBM SPSS Statistics (5th ed.). SAGE.
Gravetter, F. J., & Wallnau, L. B. (2017). Statistics for the behavioral sciences (10th ed.). Cengage Learning.
Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). Introduction to the practice of statistics (9th ed.). W. H. Freeman.
NIST/SEMATECH. (2012). e-Handbook of statistical methods. National Institute of Standards and Technology. https://doi.org/10.18434/M32189
Triola, M. F. (2018). Elementary statistics (13th ed.). Pearson.
Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.

Thank You

Author: Muhammad Hassan
Website: ResearchMethod.net

Visit ResearchMethod.net for research guides, templates, lecture slides, examples, and academic resources.